LA-UR-20-23948

Title: Uncertainty about the Uncertainty

Author(s): Makaruk, Hanna Ewa

Intended for: Virtual talk for Southern Illinois University, Carbondale, Department of Mathematics seminar series, 6/4/2020

Issued: 2020-05-29

Uncertainty about the Uncertainty

Virtual talk for Southern Illinois University, Carbondale, Department of Mathematics seminar series

Hanna Makaruk

6/4/2020

P-23, hanna_m@lanl.gov

LA-UR-20-

Abstract

A golden standard in science is to repeat an experiment a statistically significant number of times, recording data using the same set of detectors and the same data analysis methodology. In such case experimental error includes both the range of true values generated by repetitions of the experiment, and measurement uncertainty caused by the detector. They are independent. It is a huge and too frequently used simplification, to assume that one can measure multiple repetitions of an identical experiment, resulting in identical true experimental value. Repetitions, as similar is it is experimentally achievable result in a range of the true values rather than in a single value. When modern, very sensitive and well calibrated measurement systems are used, this range is not negligible, and sometimes dominates, in comparison to the measurement uncertainty. Range of true values depends on the physics of the experiment, while measurement uncertainty depends on the measurement method (properties of the detector not of the experiment).

When data from one–of –a kind experiment are analyzed, only the measurement uncertainty is reported. It gives no information about the range, in which the true values of experiment would spread if the experiment was repeated.

A frequently used approximation, that if a physical quantity is measured as a function of time, only measurement of this quantity, produces uncertainty is also in some real experiments fare to strong.

Example: In reaction history time measurement uncertainty propagated to alpha dominated under certain conditions over the flux measurement uncertainty propagated to alpha.

Reliability of a data point is in general independent from its measurement uncertainty. However, in practice reliable measurement methods frequently have high measurement uncertainty, while low reliability methods are applied to limit measurement uncertainty. Comparison of reliable data with high measurement uncertainty to not so reliable data measured with low uncertainty is discussed – in different scenarios different data analysis methods are applicable.

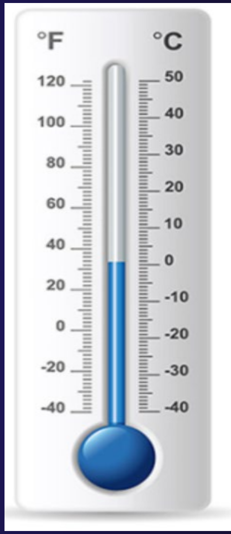
- Independent processes create independent results.
- Why does scientific community mix them up?
- Data from one-of-a-kind experiment.
- Inverse problem of the experimental data analysis.
- Unreliable or incomplete data.
- What are we talking about (glossary.)
- What kind of uncertainty do we report from a one-of-a-kind experiment, and what kind we will never be able to report.

Experimental Data Analysis is an Inverse Problem

Missing information:
low reliability or incomplete data

Experiment's Model

Inverse Problem

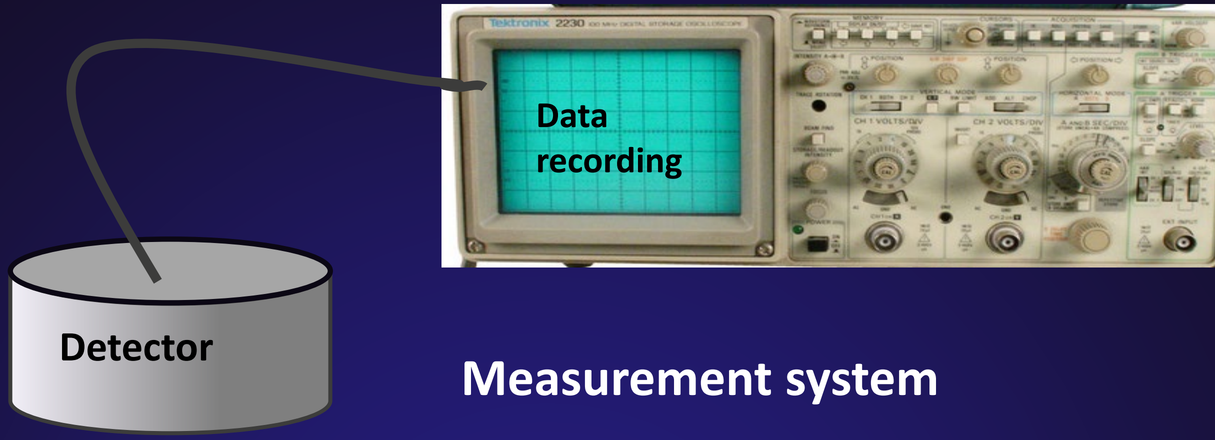


Measurement

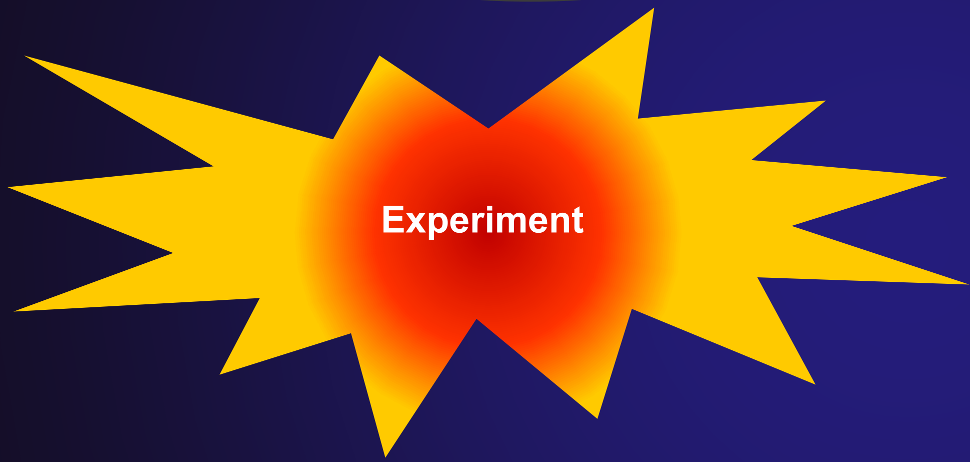
Experiment

Recorded experimental data

Experimental values



Measurement system



- Experiment determines ***Range of the true experimental value***.
- Measurement system and data analysis method determine ***Measurement uncertainty***.

Measurement uncertainty and Range of true experimental value originate from separate sources, which is why they are independent. One contains no information about the other.

Experimental measurement with no measurement uncertainty does not exist.

- When a physical quantity is measured as a function of time, the physical quantity detector (thermometer), and the time measuring device (clock) are separate sources of measurement uncertainty.

- The two resulting uncertainties are independent.

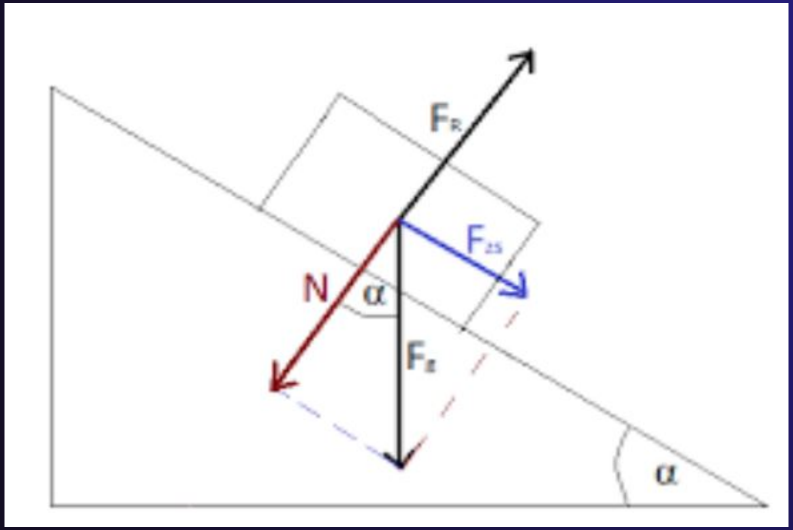
- When the quantity is plotted as a function of "exact" time, time uncertainty can contribute to the quantity uncertainty in a nonlinear way. Example – RH where alpha uncertainty originating from time measurement dominate in some regions over alpha uncertainty originating from flux measurement.

Time is always measured with uncertainty. In some experiments time measurement uncertainty is the dominant source of measurement uncertainty.

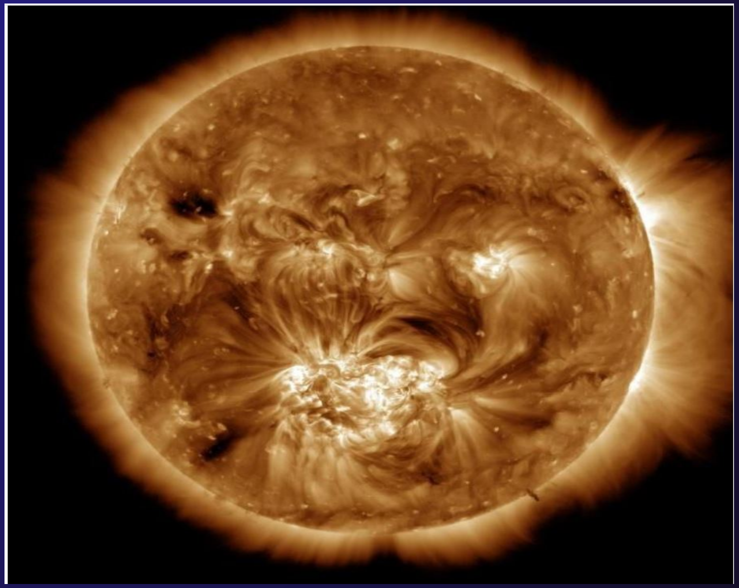
Range of true experimental values

Range of true experimental values **in which the experimental value changes, when the same experiment is repeated, if measured by a non-existing “perfect” detector.**

- **Small, unavoidable variations in experimental conditions cause a non-zero range of true experimental values.**
- **The range depends exclusively on the physics of the experiment.**
- **If no measurement is performed the range of true experimental values still exists.**



Simple experiments in classical mechanics have a narrow range of true experimental values



Sun corona (NASA)
Turbulence, a stochastic process, causes a broad range of true experimental values

"Data Uncertainty" – are we all talking about the same thing?

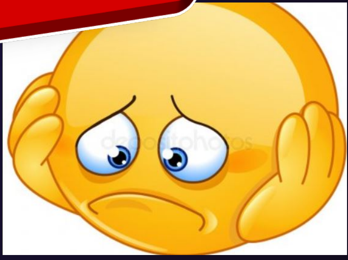
Why Scientific Community too Frequently Mixes Up Independent Objects?

- **Textbook example: experiment repeated and measured a statistically meaningful number of times.**
  - **Experimental error is calculated as a standard deviation (or similar) from a sufficiently large set of data points.**
  - **It contains everything – Range of the true experimental values and Measurement uncertainty originating from all sources.**
  - **Problem – repeating an experiment is not always possible.**

- **Early physics – seventeenth, eighteenth century**
  - **First success were in non-stochastic experiments with narrow ranges of true experimental values.**
  - **The earliest measurement methods, and detectors had limited accuracy and precision, resulting in a huge measurement uncertainty, much larger than the range of the true experimental values they measured.**

- **In quantum mechanics state of the system is changed by the process of measurement.**
  - **This does not apply to classical, macroscopic systems.**

“ ± “ is used as an upper limit on range of the true experimental values, even if “ ± “ has nothing to do with this range.

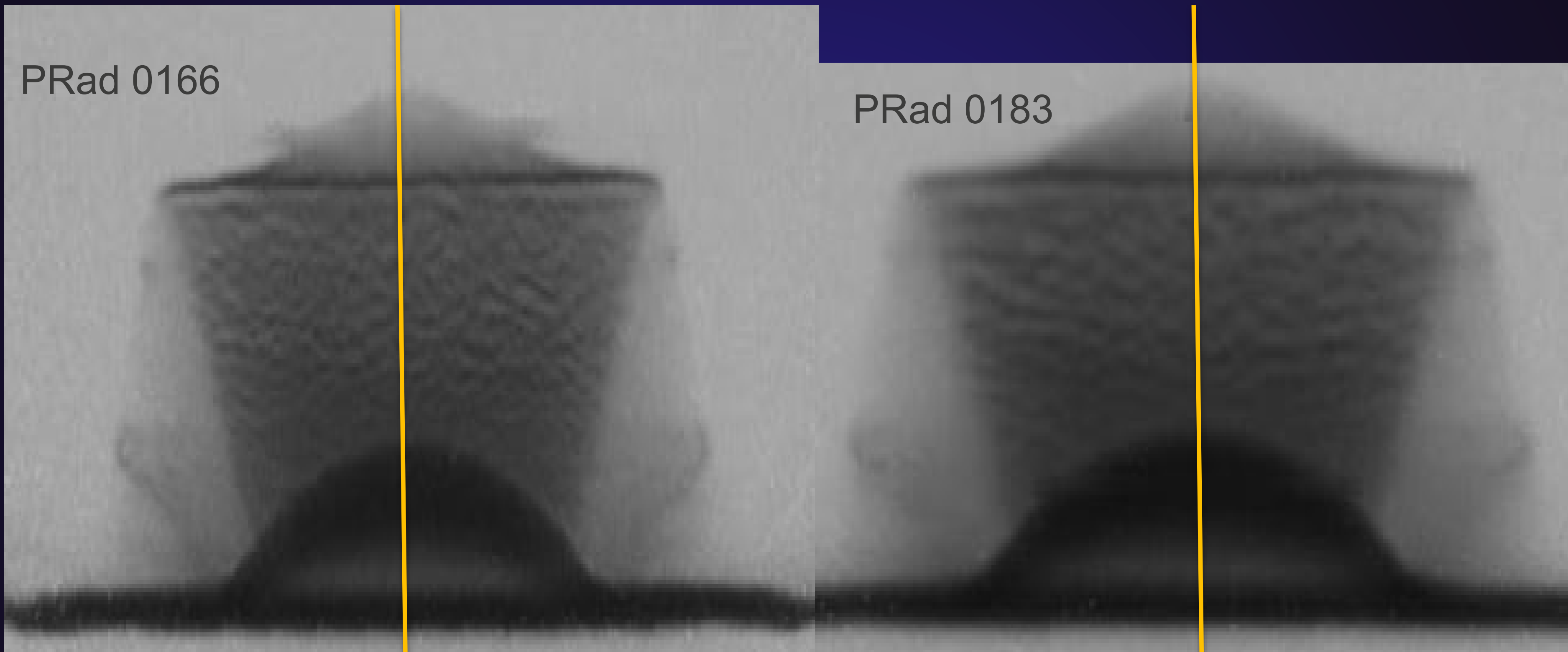


Experiment repeated statistically meaningful number of times or a set of one-of-a-kind experiments?

- **Example**: data from a month of my commuting to work, including recordings from car inside. For calculating length of the road between the office and my apartment - it is a repeated experiment. For finding what was I listening to during the trips – averaging sound tracks from the whole month would not make sense. Each trip needs to be treated as one-of-a-kind event.

The same experimental data can be analyzed, depending on the purpose, as a single experiment repeated statistically meaningful number of times, or as a set of unique one-of-a-kind experiments.

Repetition of the same experiment or two one-of-a-kind events?



Two repetitions of the experiment, as identical as experimentally achievable. PRad images of shots 0166 and 0183. The shots of melting a tin coupon 0. 3125 inch thick by explosion were as identical as experimentally achievable. The major features are preserved; yet size and shape of individual mental blobs is unique for each shot reflecting the material structure of the metal coupons.

Data analysis for one-of-a kind experiment is different than for a repeated experiment

Repeated experiment

- Golden standard in experimental science.
- Statistical methods of data analysis and experimental error calculation are well developed.
- Data sets with missing information, data affected by an unusual noise can be easily identified and excluded.
- An effect does not need to be identified to be accounted for in the experimental error.
- Total experimental error is an upper limit for the range of the true experimental values.

One-of-a-kind experiment (case study)

- In some crucial areas: from nuclear, astronomy, to medicine a lot of data is analyzed in one-of-a-kind event framework.
- Data analysis methods from repeated experiments should not be applied automatically.
- Data sets with missing information, data affected by an unusual noise create a unique problem.
- Each source of the measurement uncertainty needs to be identified, to be accounted for.
- Total experimental error can not be obtained From one-of-a-kind event analysis.

One-of-a-kind event provides no information about the range of true experimental values

Identification of all sources of measurement uncertainty

- Measurement uncertainty is a sum of uncertainties originating from multiple sources
- Each source needs to be identified and its input to the uncertainty quantified – each type of experiment needs separate account of the uncertainty sources
- There is always a possibility of overlooking an important uncertainty source – expert judgment is required.
- When data are incomplete they still need to be analyzed, this is the only data that we have.
- When data from a sufficiently large number of repetitions of the same experiment is analyzed, it is all inclusive, none of these problems exists.

Data analysis from one-of-a kind experiment creates a significantly larger challenge than analysis of an experiment repeated statistically significant number of times, and requires deep knowledge of the physics of this experiment.

Measurement uncertainty and measurement reliability

- ***Data reliability*** **– qualitative judgment by a subject matter expert, if probability of introducing a mistake during the process of obtaining, analyzing the data was high or low.**

- **Reliability and uncertainty are, in theory, independent.**

- **In practice, a valid argument to introduce a not-so-robust measurement method when a robust one exists, is when the not-so-robust method produces a result with a significantly lower uncertainty than the robust one.**

- **In case a legacy documentation is ambiguous and/or insufficient it causes low reliability in analysis of a complicated legacy measurement; analysis of a simple legacy measurements is not affected.**

- **Cross checking a low reliability/ low uncertainty result against an independent high reliability/ large uncertainty result rises reliability of the first one while preserving its low uncertainty.**

Data users assume that smaller " ± " translates to a more reliable measurement.

Old simple method: a ruler measures distance with low accuracy and precision creating a large measurement uncertainty. The measurement is very reliable – it is hard to make any mistake.

Modern, sophisticated method: interferometer measures a distance with incredibly high precision, yet this method creates multiple possibilities for obtaining incorrect result.

Low uncertainty/incomplete measurement compared to reliable, high uncertainty result

Example from 2020

1) Low uncertainty, incomplete: “This research was done on Thursday, the 20-th”

This is either 02/20/2020 or 08/20/2020 – correct, no unnecessary uncertainty added. Reporting a result with a honest either/or, brings a lot of objections from data users.

Standard reporting: May 20-th 2020 uncertainty 3 months. Formally correct, yet it is misleading.

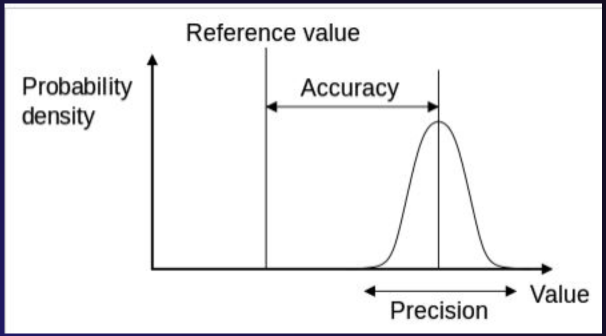
gives huge uncertainty of 6 months, even worse, many data users make unjustified Gaussian distribution probability assumption, concluding than end of May is the most probable date.

2) Adding reliable, high time uncertainty: “At Los Alamos, there was snow outside my window”

Weather pattern at Los Alamos NM, makes possible to select the exact date 02/20/2020

If information about a data point is incomplete, an independent, reliable, high uncertainty result may complete it.

Glossary



Accuracy - description of systematic errors, a measure of statistical bias; *trueness (ISO)*. (Wikipedia)

Precision – description of random errors; a measure of statistical variability (Wikipedia)

" ± " – reported in annalists' reports without description

Total experimental error – calculated be a statistical method, like a standard deviation, from data obtained in an experiment repeated sufficiently many times. It accounts for all sources of uncertainty, serves as upper limit for the range if true experimental values and for measurement uncertainty.

Range of true experimental values range, in which the experimental value changes, when the same experiment is repeated, if measured by a non-existing "perfect" detector. It depends exclusively on the experiment, not on the measurement system.

Measurement uncertainty – uncertainty created by the measurement and data analysis process.

One-of-a-kind experiment – experiment that have not been repeated; experiment that produced data analyzed separately from data produced by other, similar experiments.

Data reliability – qualitative judgment by a subject matter expert, if probability of introducing a mistake during the process of obtaining this data was high or low.

Data point error – difference between the measured value of the data point and it's true value if measured by a "perfect" detector. It is always non zero, and needs to be correctly accounted for and propagated in the data analysis methods

Perfect detector – imaginary detector that would report the measured value without any error.

Data analysis of one-of-a-kind experiment provides measurement uncertainty.

Measurement uncertainty provides no information about:

-Range of true experimental values

-Data reliability

In theory, there is no difference between theory and practice.

In practice, there is a difference.

(engineering proverb)

γ reaction history uncertainty is a measurement type uncertainty. It does not include any information about repeatability of the nuclear event.

Each data analysis of one-of-a-kind experiment is one-of-a-kind. There are multiple places, in which expert judgment is necessary.